Study
Note
2003-05

Impact of the Army Continuing Education System (ACES) on Soldier Retention and Performance: Database Development

Ani S. DiFazio
Paul J. Sticha
Human Resources Research Organization

United States Army Research Institute
for the Behavioral and Social Sciences

June 2003

U.S. Army Research Institute for the Behavioral and Social Sciences

A Directorate of the U.S. Total Army Personnel Command

ZITA M. SIMUTIS
Director

Research accomplished under contract
for the Department of the Army

Human Resources Research Organization

Technical Review by

Mark C. Young

NOTICES

DISTRIBUTION: Primary distribution of this Study Note has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: TAPC-ARI-PO, 5001 Eisenhower Ave., Alexandria, VA 22333-5600.

FINAL DISPOSITION: This Study Note may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Study Note are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE	3. DATES COVERED (from. . . to)
June 2003	Final	13 April 2001 – 30 September 2002

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Impact of the Army Continuing Education System (ACES) on Soldier Retention and Performance: Database Development	DASW01-98-D-0047/Delivery Order 0031
	5b. PROGRAM ELEMENT NUMBER 665803
6. AUTHOR(S) Ani S. DiFazio & Paul J. Sticha (Human Resources Research Organization)	5c. PROJECT NUMBER D730
	5d. TASK NUMBER 267
	5e. WORK UNIT NUMBER C01
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Human Resources Research Organization 66 Canal Center Plaza, Suite 400 Alexandria, VA 22314	8. PERFORMING ORGANIZATION REPORT NUMBER FR-02-50
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral & Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Study Note 2003-05

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Mark Young, Contracting Officer's Representative

14. ABSTRACT *(Maximum 200 words)*:

The Army Continuing Education System (ACES) provides education, training, testing, and counseling opportunities to tens of thousands of service members each year. The mission of ACES is to promote lifelong learning opportunities that sharpen the competitive edge of the Army by providing and managing quality educational programs and services. The United States Total Army Personnel Command (PERSCOM), the developers and administrators of ACES, has requested an evaluation to demonstrate the value of ACES to the Total Army. This evaluation consists of two phases. The first phase involved the development of detailed database and evaluation plans. Phase two involves the implementation of the database and evaluation plans. This report describes the data development portion of the phase two effort, which resulted in a comprehensive longitudinal evaluation database. The report presents an overview of the data structure, summarizes the activities by which the database was constructed, and discusses some of the lessons learned in the development process.

15. SUBJECT TERMS

Army Continuing Education System ACES database

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	13	Mark Young (703) 617-0334

IMPACT OF THE ARMY CONTINUING EDUCATION SYSTEM (ACES) ON SOLDIER RETENTION AND PERFORMANCE: DATABASE DEVELOPMENT

CONTENTS

INTRODUCTION

The Army Continuing Education System (ACES) provides education, training, testing, and counseling opportunities to tens of thousands of servicemembers each year. The mission of ACES is to promote lifelong learning opportunities that sharpen the competitive edge of the Army by providing and managing quality educational programs and services. The United States Total Army Personnel Command (PERSCOM), the developers and administrators of ACES, has requested an evaluation to demonstrate the value of ACES to the Total Army. This evaluation consists of two phases. The first phase involved the development of detailed database and evaluation plans. The database plan produced in phase one (DiFazio, 2000) evaluated several potential data sources and specified a methodology for producing the analytical database. After the plan was written, database development strategies were further refined to address budgetary and time constraints. Phase two involves the implementation of the database and evaluation plans. This report describes the data development portion of the phase two effort, which resulted in a comprehensive longitudinal evaluation database.

DATA STRUCTURE

PERSCOM has requested an assessment of the impact of ACES participation on soldier retention, attrition, and performance. A longitudinal database is amenable to the types of statistical analyses required to evaluate the ACES program. Longitudinal data describe relevant soldier characteristics at particular points in time. Such a database consists of a number of "blocks" of data describing attributes that can change over time, such as rank and education level. When deciding what the time interval between "blocks" should be, consideration was given to the nature of the attributes of interest to the evaluation and the frequency with which they are likely to change. A time interval that is too large should be avoided because attribute changes can be lost; a time interval that is too small will produce a database wrought with redundancy. It was determined that the attributes of interest to this evaluation are unlikely to change more frequently than once every three months. Therefore, each "block" of changing or fluid data reflects a particular calendar year quarter.[1] In addition to changing data, the evaluation plan requires data on unchanging soldier characteristics, such as demographic and accession attributes, as well as separation information.

DATABASE DEVELOPMENT ACTIVITIES

Data for the ACES evaluation were drawn from the following sources:

- Defense Manpower Data Center (DMDC) Personnel Edit File (PEF) for information on static, unchanging soldier characteristics,
- DMDC PEF for quarterly data on changing soldier attributes,
- DMDC Loss data for separation information,
- Army Education Management Information System (EDMIS) data for information on ACES participation,

[1] Data are reported as of quarter end for each of the 24 quarters in FY 96 through FY 01.

- Servicemembers Opportunity Colleges Army Degree (SOCAD) program data for SOCAD 2- and 4-year contracts,
- Selective reenlistment bonus (SRB) data from Army SRB directives, and
- Performance and participation data from another study conducted for the Army Research Institute (ARI), called the NCO21 Validation Study.

HumRRO performed all data processing tasks using the Statistical Analysis System (SAS).

DMDC Data

The variables that were included in the written request sent to DMDC for PEF and Loss data are presented in Appendix A.[2] In addition to this written request, HumRRO staff were in frequent communication with relevant DMDC staff regarding the request throughout the data-gathering process.

DMDC began their work by identifying the relevant population of non-prior service accessions from October 1995 through September 1998 from Military Enlistment Processing Command (MEPCOM) data. This resulted in a population of 203,630 soldiers. Next, DMDC extracted the requested static PEF variables for these soldiers and transmitted the data to HumRRO as a text file. The static data consisted of demographic information (e.g., gender, race) and data at the time of accession. An additional text file was provided to HumRRO containing Montgomery GI Bill (MGIB) data. HumRRO read these data, merged the MGIB with the static data, performed diagnostic checks, and made the required edits to clean the data.

DMDC further provided 24 text files containing changing data from the PEF for each quarter in the period October 1995 through September 2001. Each file contained data on the 203,630 soldiers in the total population, regardless of whether the soldier was in service during the quarter or not.[3] HumRRO read these data, deleted records of soldiers who were not present during the quarter, performed diagnostic checks, made the required edits to clean the data, and merged the 24 files to create a single record consisting of quarterly data for each soldier.

DMDC also extracted separation data from the DMDC Loss Files and transmitted the data to HumRRO as a text file. The separation data contained at least 1 record for each of the 203,630 soldiers in the population, regardless of whether a loss had actually occurred.[4] HumRRO read these data, deleted records of non-separated soldiers, performed diagnostic checks, and edited and cleaned the data. There were no more than three separations (which include reenlistments) per soldier for the population at hand. A database was produced consisting of a single record per separated soldier describing up to three separations.

---

[2] The request did not distinguish the source of the requested data. In fact, DMDC and HumRRO investigated several potential sources of the data.

[3] The only non-missing fields for those soldiers who were not in service during the quarter in question were Social Security Number (SSN) and accession year and month.

[4] Except for the SSN field, records of soldiers who had never separated consisted exclusively of missing data.

It should be noted that, despite our attempts, we were unable to obtain detailed definitions of several DMDC date variables.[5] These definitions would have enhanced our understanding of the data in general and the relationships between the numerous date variables maintained by DMDC in particular.

EDMIS Data

The evaluation plan specified the population of interest as only those of the 203,630 soldiers in the accession cohort who were exclusively in EDMIS-operational facilities. Not all of the approximately 114 facilities with EDMIS currently installed have had the system operational long enough to include in an evaluation containing October 1995 through September 1998 accessions. Thirty-six facilities[6] that became operational on or before October 1997 were selected: 27 of these were operational on or before October 1995, 3 on or before October 1996, and 6 on or before October 1997. Using the DMDC variable "Duty Base Identifier" from the DMDC "changing" data, 51,764 soldiers were identified as serving only in the selected EDMIS-operational facilities. A list of selected facilities, the dates that EDMIS became operational, and the duty base identifier codes is presented in Appendix B.

The SSNs for the 51,764 soldiers in the evaluation population were sent to PERSCOM for matching to EDMIS system data from the selected facilities. Data from six EDMIS tables thought to be potentially useful to the evaluation analyses were extracted and sent to HumRRO as individual Microsoft® Access tables. After review of the data and numerous discussions with PERSCOM staff, three tables[7] were eliminated from additional processing due to unreliability or inconsistent usage at the EDMIS sites. The evaluation database includes data from the following EDMIS tables:

- COLLTAEN – All College Enrollments
- SEPENR – Individual SEP Enrollments
- IND_ALC_TBL – Individual Academic Learning Center Visits

HumRRO converted the Access tables into SAS data. Each EDMIS file contained as many records for an individual soldier as there were COLLTAEN, SEPENR, and/or IND_ALC_TBL "events" for that soldier. The participation date in each record was evaluated, so that the information on that record could be assigned to the correct quarter. Diagnostics were run on each file and data were re-coded, cleaned or eliminated based on recommendations from PERSCOM staff. Next, relevant EDMIS participation summary indices were created within each of the 24 quarters in the evaluation. The result of these activities was a single record per soldier of cleaned data describing EDMIS participation in each of the 24 quarters of interest.

Two- and four- year SOCAD contract data through September 2001 for the EDMIS population were extracted by Servicemembers Opportunity Colleges personnel. HumRRO read

---

[5] These variables are: Date of Accession, Initial Entry Calendar Date, Longevity Pay Service Base Calendar Date, And Active Federal Military Service Base Calendar Date.

[6] Unique Duty Base Identifiers rather than location names are the basis of this count.

[7] The three eliminated tables were TABEFIL, ARMYPERS, and MILGAIN.

these two text files and created a SAS database consisting of SSN and whether the soldier had contracted a 2- and/or 4- year SOCAD agreement.

SRB Multiplier Data

The SRB multiplier is an important control variable to be considered when analyzing effects of ACES participation on reenlistment. Although administrative data sources record the SRB multiplier for soldiers who reenlist, a complete record of the SRB multipliers for which each soldier is eligible must be obtained from the memoranda that document Army reenlistment bonus policy. A SAS file containing over 15,000 SRB directives in effect since October 1995 was assembled. These data include the directive MOS and effective date, and, where applicable, the grade, location, additional skill identifier (ASI), and special qualification indicator (SQI). The SRB directive in effect for each of the 51,764 EDMIS soldiers in each of the 24 quarters was identified based on soldier MOS, grade, location, ASI, and SQI. The appropriate SRB multiplier was then assigned to every soldier in each of the 24 quarters of interest.[8]

Final Longitudinal Database

The individual files described above were merged by SSN to form a single record for each evaluation population soldier. To facilitate identification of variable data sources, the following variable name prefixes were assigned:

- C = DMDC changing data
- D = SOCAD data (2- and 4- year contracts)
- E = EDMIS COLLTAEN data
- I = EDMIS IND_ALC_TBL data
- O = DMDC loss data
- Q = SRB multiplier data
- S = EDMIS SEPENR data
- U = DMDC unchanging data

The variable names of data that change were also assigned a three-character suffix indicating calendar year quarter and year. For example, a variable name ending in "201" describes data as of the second calendar year quarter (April-June) of the year 2001. In addition to this informative variable nomenclature, each variable (except SSN) has a label that describes the data element and indicates the source of the variable.

As discussed above, each individual file that contributed to the final merged data was subjected to intra-file consistency and accuracy diagnostic checks. The types of questions addressed by these diagnostic checks were as follows:

- Are there variable value outliers?

[8] If there was more than one directive in effect for a soldier during a quarter, then the one with the largest SRB multiplier was used.

- Do dates make sense given the evaluation population accession cohort?
- Do dates make sense compared to each other (e.g., does accession date always precede separation date)?
- Are there any cases where education level decreases over time?
- Are grade changes always accompanied by changes in date of rank?

The final merged data were subjected to series of inter-file diagnostic checks. These checks were performed to assess the logical relationship between variables from different sources. The types of questions addressed by inter-component file consistency checks performed were as follows:

- Are the EDMIS sites the same between COLLTAEN, SEPENR, and IND_ALC_TBL in each quarter?
- Are the Duty Facility Identifiers from DMDC and the EDMIS site locations the same for each quarter?
- Is there changing DMDC data after there has been a non-reenlistment loss?
- Do we see EDMIS participation data after there is a non-reenlistment loss?

HumRRO reviewed output from all inter- and intra-component file diagnostic checks and edited or deleted data programmatically to ensure data consistency and accuracy. The final data consists of 43,831 records and 1,830 variables[9].

NCO21/ACES Performance Evaluation Data

Soldier background information, supervisor performance ratings, and self-reported ACES usage data already developed by HumRRO for ARI's NCO21 validation study were merged with updated Enlisted Master File (EMF) data to serve as the analysis file for ACES performance analyses. To facilitate identification of variable data sources, the following variable name prefixes were assigned to variables:

- B = NCO21 soldier background information
- DZ = DMDC accession data
- F = NCO21 supervisory future ratings
- FZ = EMF data as of 6/01 collected as part of the ACES database project
- J = source data mapping variables
- MZ = EMF data as of 6/01 collected as part of the NCO21 project
- O = NCO21 supervisory observed ratings
- P = NCO21 Personnel File Form-21 (PFF) data

These data form a separate database from the main evaluation data described above. Of the 1893 soldiers in the NCO21 data, 114 also appear in the main database[10].

[9] Contents of the longitudinal database can be obtained from ARI by request.
[10] Contents of the NCO21/ACES database can also be obtained from ARI by request.

PROBLEMS AND LESSONS LEARNED

The fact that there were problems in the development of the database was not surprising. Our experience with other database development projects has led us to anticipate a certain level of error and inconsistency in military administrative data. Although the existence of problems was not unanticipated, the extent and sources of data errors were sometimes surprising. In addition, the problems were multiplied in magnitude because of the requirement to combine information from different sources.

Some of the problems that proved to be most difficult to solve came from inconsistencies between data from different sources. These inconsistencies were most evident in date variables. For example, cases were found with changing DMDC data subsequent to a non-reenlistment loss where the date of last enlistment did not change. In general, we could not resolve these discrepancies and consequently deleted inconsistent records.

One of the most important facilitators of success in the database development is the existence of a point of contact (POC) at the agency that maintains the database to answer questions, advise about data quality, and clarify any ambiguities. Our experience with this effort indicates that an effective and engaged POC can greatly increase the speed at which the effort progresses, as well as reduce the blind alleys that must be investigated.

REFERENCE

DiFazio, A.S. (September 2000). Impact of the Army Continuing Education System (ACES) on soldier retention and performance: A database development plan. Alexandria, VA: Human Resources Research Organization.

APPENDIX A
VARIABLES INCLUDED IN DMDC DATA REQUEST

SEQ #	Item Name
1.	Person Social Security Number Identifier
4	Uniformed Service Pay Grade Code *
5	Pay Grade Modifier Code *
6	Uniformed Service Rank Effective Calendar Date *
7	Assigned Unit Identification Code *
8	Duty Unit Location Country Code *
9	Duty Unit Location US State Alpha Code *
10	Duty Unit Location US Postal Region ZIP Identifier *
13	Duty Base Facility Identifier *
14	Uniformed Service Initial Entry Calendar Date
15	Military Longevity Pay Service Base Calendar Date
16	Military Longevity Pay Service Years Quantity *
17	Active Federal Military Service Months Quantity *
18	Active Federal Military Service Years Quantity *
19	Primary DoD Occupation Code *
20	Primary Service Occupation Code *
21	Duty DoD Occupation Code *
22	Duty Service Occupation Code *
23	Secondary DoD Occupation Code *
24	Secondary Service Occupation Code *
25	Person Birth Calendar Date
26	Person Age Quantity *
27	Person Sex Code
28	Race Code
30	Race Ethnic Code
31	AFQT Percentile Quantity
32	Dependents Quantity *
33	Marital Status Code *
35	Joint Service Spouse Data Source Code *
36	Joint Service Spouse Service Branch Classification Code *
37	Joint Service Spouse Uniformed Service Organization Component Code *
38	Joint Service Spouse Pay Plan Grade Code *
40	Enlisted Active Service Agreement Begin Calendar Date *
41	Accession Program Source Code
43	Educational Level Code *
45	Home of Record Country Code *
46	Home of Record US State Alpha Code *
47	Legal Residence US State Alpha Code *
53	Person Surname Text
58	Active Federal Military Service Base Calendar Date
59	Collocated Dependents Type Code *
60	Collocated Dependents Quantity *
61	Command Sponsored Dependents Quantity *
62	Assigned Unit Major Command Code *
63	Active Duty Strength Accounting Code *
64	Active Duty Involuntary Retention Reason Code *
66	US Citizenship Status Code *

An asterisk indicates time-varying variables.

SEQ #	Item Name
68	Enlisted Active Service Agreement Duration Years Quantity *
69	Permanent Duty Station Arrival Calendar Date *
70	Permanent Duty Station Departure Calendar Date *
71	Enlisted Active Service Projected End Calendar Date *
73	Enlisted Career Status Code *
74	Selective Reenlistment Bonus Multiplier *
75	Accession Home of Record US Postal Region ZIP Identifier
76	Accession Marital Status Code
77	Accession Prior Service Military Indicator Code
78	Accession Enlisted Active Service Agreement Duration Years Quantity
79	Accession AFQT Category Code
80	Accession Training Service Occupation Code
81	Accession AFQT Percentile Score Quantity
83	Accession Educational Designator Code
84	Separation Program Designator Code/Accession Designator Code *
85	Interservice Separation Code *
86	Personnel Transaction Source Code *
87	Personnel Transaction Type Code *
88	Transaction Effective Calendar Date *
89	Reenlistment Eligibility Code *
90	Military Service Characterization Code *
92	Educational Discipline Code *
93	File Calendar Date *
94	Personnel Transaction Unreconciled Status Months Quantity *
95	Loss Category Code *
97	Expiration of Enlistment on Active Duty Calendar Date/Expected Active Duty Loss *

An asterisk indicates time-varying variables.

APPENDIX B
LIST OF INCLUDED EDMIS SITES

Facility Location	Date EDMIS Operational	DMDC Duty Base Identifier Code
EDMIS Sites Operational on or before October 1995		
Redstone Arsenal, AL	October 1995	01013A
Fort Bragg, NC	June 1995	37008A
Fort Campbell, KY	March 1994	21001A
Fort Drum, NY	1994	36007A
Fort Hood, TX	1995	48012A
Fort Lewis, WA	January 1995	53007A
Fort Polk, LA	1994	22006A
Fort Belvoir, VA	1994	51010A
Fort Meade, MD	1993	24006A
Fort Myer, VA	1994	51014A
Pentagon:		
Dept Air Force Activities	1994	51007F
Dept Army Activities	1994	51008A
Dept Navy Activities	1994	51009N
Carlisle Barracks, PA	April 1992	42002A
Fort Bliss, TX	April 1993	48011A
Fort Eustis, VA	March 1995	51011A
Fort Gordon, GA	June 1992	13006A
Fort Jackson, SC	October 1994	45006A
Fort Lee, VA	1989	51012A
Fort Monroe, VA	March 1995	51013A
Fort Rucker, AL	November 1992	01007A
Fort Sill, OK	June 1994	40003A
Fort Story, VA	March 1995	51015A
Fort Richardson, AK	Summer 1994	02007A
Fort Shafter, HI	October 1994	15003A
Fort Wainwright, AK	Summer 1993	02006A
Schofield Barracks. HI	October 1994	15006N
Tripler Army Medical Center, HI	October 1994	15003A (same as Shafter)
Additional EDMIS Sites Operational on or before October 1996		
Fort Carson, CO	January 1996	08004A
Walter Reed Medical Center, D.C.	1996	11007A
Fort Leavenworth, KS	1996	20004A
Additional EDMIS Sites Operational on or before October 1997		
Fort McPherson, GA	May 1997	13007A, 13005A
Fort Stewart, GA	October 1997	13008A
Hunter Army Airfield	October 1997	13008A (same as Stewart)
Fort Benning, GA	October 1997	13004A
Fort Huachuca, AZ	Summer 1997	04003A
Fort Knox, KY	December 1996	21002A